System for charging the use of a packet-based telecommunication network.

BACKGROUND OF THE INVENTION

The invention is related to a system for charging, ~~the use of~~ in a packet-based telecommunication network, such as an ATM or IP oriented network, the packet load per connection.

a. ATM (Asynchronous Transfer Mode)

ATM is a relatively new network technique for supplying connections with very different characteristics in a uniform manner. For each ATM connection has to be paid. Preferably, the amount charged for the connection reflects both the performance delivered by the network and the performance experienced by the customer. In a telephone network something like this is achieved by letting the amount to be dependent on the duration of the connection (calculated in seconds, minutes or other units, such as "ticks") and on the distance covered. A telephone call, however, is almost always a case of a connection with a fixed capacity (e.g. 64 kbit/s with ISDN). In contrast, ATM is much more flexible and there are more connection parameters and variables than with PSTN or ISDN based connections.

- With ATM connections can be established with varying capacity (some kbit/s up to hundreds of Mbit/s).
- With ATM connections can be established with varying network garantuees concerning cell loss, cell delay, cell delay variance and throughput by choosing an "ATM Transfer Capability" (ATC) [2] and Quality of Service class (QoS class) [3]. The amount charged for an ATM connection should preferably also reflect this additional flexibility, which is the subject of the present invention. An important aspect in determining the manner of charging for ATM connections is the direction (incentive) that the charging gives to the manner of network usage. In a traditional telephone network the time-related charge usually assures that the customer does not unnecessarily occupy the the connection. In a traditional data network usually a volume rate is used, so that there is an incentive not to burden the network with unnecessary traffic. Another example is the application of an off-peak rate with a view of shifting a part of the network traffic to periods outside the peak hours so that the network can be dimensioned smaller and thereby is less expensive.

The present state of the art usually envisages to base the charge of an ATM connection on two variables, i.e. a time component, the duration of time of the connection (session), and a volume component, the total number of ATM cells transmitted and/or received during the connection [1]. Both variables can easily be measured, registered and

processed into a charge during the connection. On processing of the values of the time component the price per unit can depend on various quantities. Examples of such quantities are the distance covered and time of the day or of the week, comparable with the usual charging for telephone. Examples of other quantities are the ATM parameters such as the Peak Cell Rate of the connection etc.
There are different ATCs standardised in ITU-T recommendation I.371. Hereafter it is indicated what are the limitations for some of the ATCs of choosing a charge which is based on the time and volume component as described in the above.
A charge which is exclusively based on the time and volume component as set out in the above, has as a consequence that only the total connection duration and only the total number of cells during the duration of the connection play a role in the charge. For these quantities (and thereby for the charge) it does not matter whether all cells are offered evenly over the connection duration ("Constant Bit Rate") or are concentrated in one or more bursts of cells ("Variable Bit Rate"). For the network it is advantageous if the cells offered are spread as much as possible. ATM connections that use SBR (Statistical Bit Rate) are characterised by two additional parameters, the Sustainable Cell Rate (SCR) and the Maximum Burst Size (MBS). The essence of the situation set out above does not change, however: the user experiences no incentive to spread the cells as much evenly as possible, while this is beneficial for the total network capacity and consequentially is pursued by the network operator. The question is thus in which way the user can be urged to offer the traffic as evenly as possible. In other words, a method is needed to urge the network user via the charging mechanism to offer the traffic in a less bursty way. If in an ATM network use is made of an ABR (Avaialable Bit Rate) control mechanism, the network dynamically assigns capacity to each connection. It can however occur that the network assigns capacity to a connection but that the user does not use or not comletely use that capacity., e.g. if the user sends less cells than the assigned capacity allows. With the present charging mechanisms (based on a total time and a total volume component in the charge), leaving capacity assigned by the network unused leads to a lower charge. There is no incentive to use the assigned capacity indeed and there is no incentive to gear the capacity to the actual current need.
b. IP (Internet Protocol)
The IP is a connectionless packet switched technique which is used for the Internet. Current IP networks exclusively supply a so-called best-effort service. The networks commits itself to make an effort to deliver the packet (datagramme) at the destination but no guarantee is

given; the packet can be lost in case of a congestion. It is customary to charge access to the Internet only, e.g. by a fixed amount per month (flat rate) or by a fixed rate per unit of time (hour) that the user is logged onto his Internet Sevice Provider. In this type of charging there is no relation with the amount of data wich a user asks or offers.

Because IP is a connectionless technique, there is no matter of "connections" in the same sense as with a telephone or an ATM connection. The aspect of time is therefore inherently unsuitable to serve as measure for network load. To be able to relate the use of an IP network to the network load caused by the network user, an other quantity has to be used.

The amount of data can in IP be expressed in different units, such as the number of datagrammes per unit of time and the number of bytes (or bits) per unit of time of which the datagrammes consist.

Guaranteed IP services

Recently, work is being done on an extension of the services which an IP network can offer. The aim is, next to the best-effort service described in the above, to also enable an IP network to give guarantees for the throughput and for the delay experienced in the network, comparable with the possibilities that an ATM network offers. The standardisation of these new services with guarantees is still in an early stage.

One of the proposed possibilities is to use reservations, e.g. with the protocol RSVP [4]. In that case it is desirable that the extent of the reservation requested or made and the duration of the reservation is expressed in the charge.

In an other proposed approach, some bits in the IP header are used to indicate to which service class the IP packet belongs, e.g. "best-effort" or "guaranteed with short delay". In that case it is desirable that the indication of the service class also is expressed in the charge.

All considerations as mentioned above for ATM/DBR and ATM/SBR are mutatis mutandis also applicable on these new IP services with guarantees.

SUMMARY OF THE INVENTION

The invention provides a charging system in which the charging gives more direction to an efficient network use. To this end, the invention proposes not to measure and charge the total numbver of ~~data unit~~packets (cells, IP datagrammes, bytes in IP datagrammes) during the whole connection (session), but to subdivide a connection in shorter or longer measuring periods, to measure the number of ~~data unit~~packets during such measurement periods and base the charging on

that. The invention comprises hereunto a measurement device for measuring the number of ~~data unit~~packets received and/or transmitted during a set period of time, shorter than the time during which said telecommunication connection is open or active. In stead of measuring the number of ~~data unit~~packets over a fixed period, as known from reference 5 (Lindberger), it is conversely also possible to measure the duration of time between the reception or transmission of a specific number of ~~data unit~~packets. This has a substantial advantage over the method known from Lindberger. Especially when the time lapse t between two or a relative small number N of packets of the same (virtual) channel is measured, for each connection channel --some channels having a small number of packets per time unit, while other channels have a great number of packets per time unit- charging according to the ratio N/t will cause, for both kinds of channels, a price incentive for the channel users to use their channels as "smooth" as possible.

Furthermore, the invention comprises a calculation device for calculating for each set or measured period of time the number of ~~data unit~~packets per unit of time and offering that calculation result a billing system. The calculation device calculates thus per û shorter or longer û period the real ~~data unit~~packets / time ratio, whereby the billing follows the actual network load more accurately. Thus, for the user an incentive can be created not to offer the data in bursts but more evenly and thereby contributing to a more efficient network use. The measuring period can be equal to the interarrival time of two consecutive cells of a same connection. The rate over the period from ti til and including ti+1 is then equal to 1/(ti+1 û ti), in other words, the inverse of the difference of the arrival and send times of two consecutive cells. The measuring period can also be longer, e.g. the time between cell number i and cell number i+n, where n>1. The measuring period can also be a set period, e.g. 100 ms. It will be appreciated that the shorter the measuring period is, the more accurate the measurements are, but also the larger the calculation capacity of the charging computer has to be. Also, it requires transmission traffic between the charging measurement points and the charging computer.

Registering all measured information for all connections can lead to a large amount of data between the registration device and the billing system. A decrease of the amount of data can be achieved by aggregating the data in an aggregation device and transmitting the aggregated data to the billing system.

If a data stream is controlled by the ABR mechanism in ATM, the cell rate assigned (dynamically) by the network is mentioned in the ECR

field (Explicit Cell Rate) of so-called backward RM cells. To base the charge, apart from the real transmitted cell rate per measuring period, as proposed in the preceding, also on the capacity assigned by the network, the system can be extended by a device which reads the value that is written in the ECR field. In the same way, the cell rate desired by the user, mentioned in the ECR field of the so-called forward RM cells, can be detected and processed. In that way, the charging is based both on the capacity desire of the user and on the capacity which the network assigns to the user. A comparable function is accomplished in an IP network by reading and registering the size of the reservations desired or made from the reservation messages (e.g. RESV messages of RSVP [4]), or by reading and registering the priority indication in the header of the IP datagramme, and adjusting the charge accordingly.

DESCRIPTION OF THE FIGURE

Figure 1 shows an exemplary embodiment of the invention.

ATM: Asynchronous Transfer Mode

A physical communication line 1 transports ATM cells. The cells can belong to different virtual connections (channels, paths). A measuring device 2 detects from the header of an arriving cell the virtual connection to which the cell belongs. In the measuring device 2 for each (virtual) connection a counter reading is kept up to date with the number of arriving cells. A clock generator 3 generates periodical clock pulses. A calculation device 4 calculates per connection the ratio between the number of arrived cells and the number of clock pulses and passes this ratio on to a billing system 5. According to the invention, said ratio is not calculated over the entire time that a connection is active but over smaller periods. There are therein two possibilities, viz. (per connection) starting from a fixed measuring period T and counting the number n of cells arriving in that period, wherein the ratio r = n/T, or starting from a fixed number of cells N and measuring the time t which is needed for the arriving of those cells, wherein r = N/t.

In order to charge a bursty supply of cells more heavily than an even supply, the fixed periods of measurement T or the fixed number of cells N is chosen in such a way that the ratio n/T or N/t is high in traffic with a (temporary) "burst" character. To create an incentive for an even traffic supply, measurement periods in which the calue of r is high are charged more heavily than periods with a low value of r. In order to let the billing, in a connection which uses ABR, also be dependent on the capacity desired by the user, that value r1 is read by device 2 from passing RM (Resource Management) cells and passed on, through the calculation device 4, to the billing system 5. In the same

way a value r2 is extracted, via a return connection 1', in which measuring device 2', a clock generator 3' and a calculation device 4 are included, from the "backward" RM cells, which is an indication for the maximum cell rate which is (dynamically) assigned by the network to the user. That value too, is passed on to the billing system and is included, as well as the value r1, in the price to be charged to the user.

For use in an IP network (where the packet size may vary) the measuring device (2) is moreover able to measure and register the size of the packet (datagramme). The registration includes then the number of IP datagrammes and the cumulative number of bytes in those datagrammes. A calculation device 4 calculates per connection the ration between the number of arrived datagrammes/bytes and the number of clock pulses and passes both ratios on to a billing system 5. In the same way, via a return connection 1', in which a measuring device 2', a clock generator 3' and a calculation device 4 are included, a value r2 is in the opposite direction extracted from the reservation messages, which is an indication for the reservation promised by the network. That value too is passed on to the billing system and is included, as well as the value r1, in the price to be charged to the user. Instead of relating to the capcity, the parameters r1 and r2 can also relate to the (requested or assigned) priority of the datagrammes.

Optionally, as the figure shows, an aggregation device 6 can be added. This device aggregates the periodically generated data from the calculation devices 4 and 4', so that the task of the billing system 5 is relieved and the quantity of billing data to be transported is reduced.

REFERENCES

[1] Marion Raffali-Schreinemachers et al: "Charging and billing issues in high speed heterogeneous networking environments"; Proceedings InterWorking '96; October 1-3, 1996; Nara, Japan; pp. 97-108.

[2] ITU-T I.371: "Traffic Control and Congestion Control in B-ISDN"; ITU-T recommendation I.371 (08/96); Geneva, August 1996.

[3] ITU-T I.356: "B-ISDN ATM layer cell transfer performance"; ITU-T recommendation I.356 (10/96); Geneva, October 1996.

[4] Braden, R. (Ed.), et al: "Resource ReSerVation Protocol (RSVP) Version 1, Functional Specification"; Internet Engineering Task Force, RFC 2205, September 1997.

[5] Lindberger, K, "Analytical methoeds for the traffical problems etc.", Teletraffic and datatraffic etc., Copenhagen, june 19-26, 1991